Fault Management Metrics

Dr. Stephen B. Johnson
President, Dependable System Technologies

Dr. Sudipto Ghoshal
Vice President, Engineering, Qualtech Systems, Incorporated

Dr. Deepak Haste
Director, Engineering, Qualtech Systems, Incorporated

Dr. Craig Moore
NASA Marshall Space Flight Center
EV43 Integrated System Health Management and Automation Branch

For AIAA SciTech Conference
Grapevine, Texas
11 January 2017

Goal of Paper

- **Present metrics by which the effectiveness of Fault Management can be estimated**
  - System Health Management = the capabilities of a system that preserve the system's ability to function as intended
  - Fault Management = operational subset of System Health Management
  - FM = the operational capabilities of a system that preserve the system's ability to function as intended

- **Provide example calculation with generic "fake" data**
  - Deep space probe example
  - Simplified set of scenarios and fault management, not comprehensive
  - Won't address probabilistic distributions and uncertainties, though those are very important; will use "mean point estimates"

- **The method described here has been successfully applied to the NASA Space Launch System (SLS) program for "aborts".**
  - Abort = capability to enable the crew to escape from the launch vehicle hazard and return back to Earth

FM Theory => FM Metrics

- **FM is implemented as a set of “meta-control loops” that aim to restore the system to a state that is controllable by nominal (passive and/or active) control systems**
  - Usually the regular (passive or active) control system has been compromised because (for active control) its sensors, processing, or actuators are compromised, or (for passive control) the design margins have eroded to zero or negative.

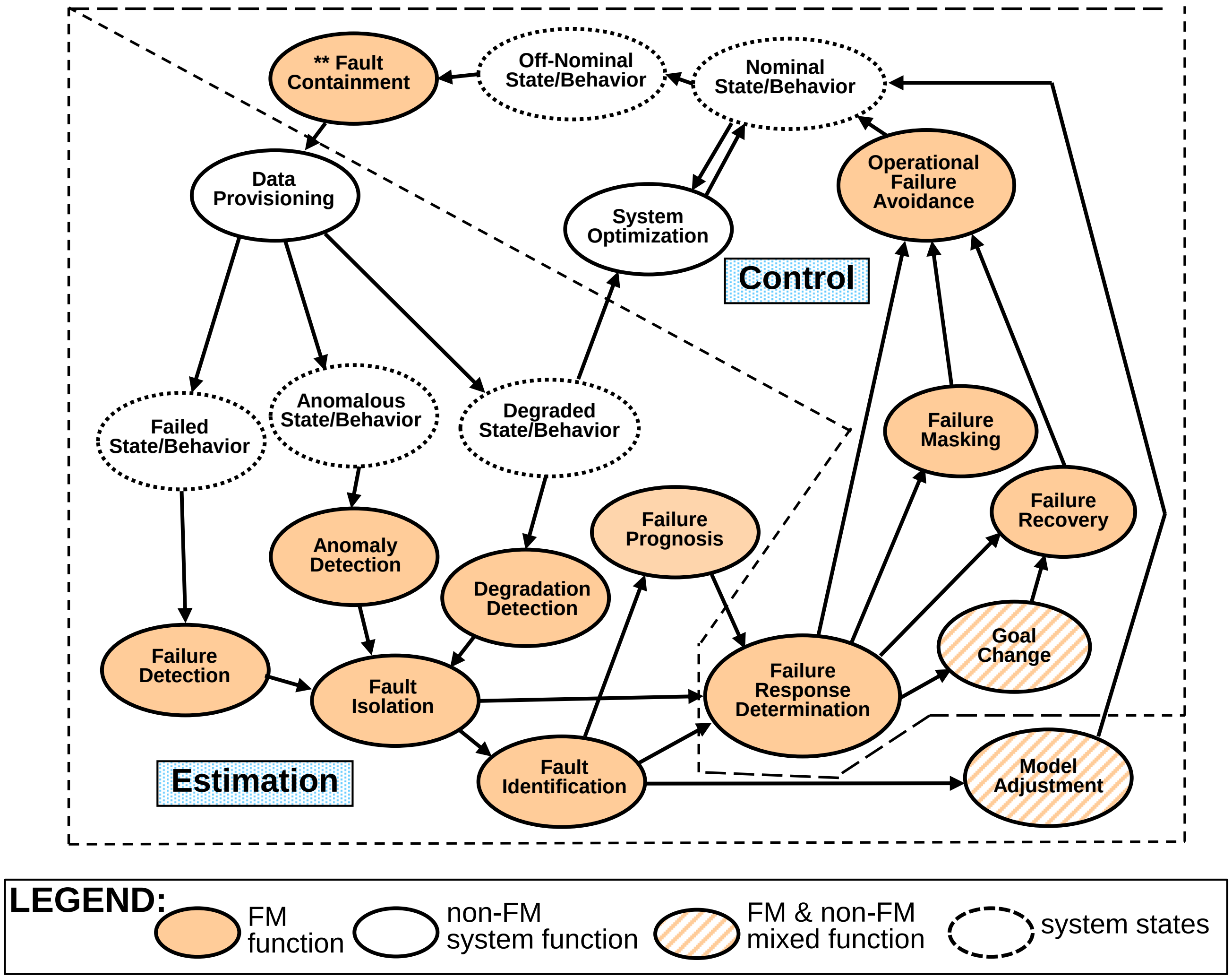
- **Each FM Control Loop (FMCL): failure detection, fault isolation, decision, and response**
  - Variants include different detection types (anomalies or degradations), prognostics, failure identification, and different response types (recovery, goal change, operational failure prevention).

- **Control theory applies**
  - State estimation and control = failure detection/isolation and failure response decision/execution

- **FM Metrics based on this extension of control theory, and use state estimation and state control for each FMCL**

Fault Management Functions as Control Loop

** Function may exist anywhere in FM loop



** Fault Containment

Off-Nominal State/Behavior

Nominal State/Behavior

Data Provisioning

System Optimization

Operational Failure Avoidance

Control

Failed State/Behavior

Anomalous State/Behavior

Degraded State/Behavior

Failure Masking

Failure Recovery

Failure Prognosis

Anomaly Detection

Degradation Detection

Failure Detection

Fault Isolation

Failure Response Determination

Goal Change

Estimation

Fault Identification

Model Adjustment

LEGEND:

FM function

non-FM system function

FM & non-FM mixed function

system states

State Estimation & State Control Metrics

♦ State Estimation Metrics

- Truth Table or “Confusion Matrix”
- For Failure Detection:
  - True Positive (TP): A correct estimate that a failure exists, when it really does exist.
  - True Negative (TN): A correct estimate that a failure does not exist, when it really does not exist.
  - False Positive (FP): An incorrect estimate that a failure exists, when it really does not exist.
  - False Negative (FN): An incorrect estimate that a failure does not exist, when it really does exist.
- TP/TN/FP/FN *ALWAYS* used for state estimation metrics, though defined somewhat differently for diagnostics and prognostics

♦ State Control Metrics

- Correctness of response (set) selected (can be a series or set of responses)
- Effectiveness of response execution
- Timing is an essential feature of state control
  - Race Condition of failure effect propagation vs. FMCL latency
  - FMCL must execute and complete before failure effects propagate to the “Critical Failure Effect”
  - Timing factor related to control loop “characteristic time” and relation to system physics

Failure Scenarios

- FM metrics are calculated for each relevant FMCL in each credible failure scenario

- To get total system metrics, the scenario-level estimates can be summed to get system-level estimates

- Example: Science-gathering planetary spacecraft (not lander) - loss of attitude control
  - Several possible causes, including failures of sensors, controller, actuators, with many failure modes within each of them.
  - Different behaviors and criticality during cruise vs. separation vs. orbit insertion vs. science-gathering (assume spacecraft gathering at a planet/asteroid)
  - Scenarios defined by the failure effect (loss of attitude control) and system configuration and/or mission phase in which the failure occurs
  - FMCLs: Could be detected by attitude control detection or direct measurements of actuator performance. Responses include switching sensors, controllers, actuators, safing, etc.

- For SLS, there are currently ~16,000 failure modes, and these are compressed to ~1,300 scenarios because the scenarios are defined by "Intermediate failure effects" to which several or many failure modes contribute. Scenarios often assessed in groups, for SLS ~200 groups.

Example: Failure Scenarios & FMCLs

Table 1. Example Failure Scenarios and Estimated Risk Values. Total Risk in these scenarios: 2.14E-3.

	Cruise (C)	Orbit Insertion (O)	Science Gathering (S)
Loss Computing (GLC)	1E-4	1E-5	3E-4
Loss of Thermal (LTC)	1E-5	Not Credible	2E-5
Loss of Att Ctl (LAC)	2E-4	3E-4	5E-4
Propellant Leak (PRP)	3E-4	Not Credible	4E-4

- **The FMCLs to be assessed against these scenarios are as follows:**
  1) **Watchdog Timeout; Computer Reboot and Safing (WT-Rbt&Sf)**
  2) **Mission Operations Thermal Trending; Attitude Profile Redesign and Power On Extra Components (Trnd-Att&PWR)**
  3) **Attitude Control Failure Detection; Sequenced Switch to Redundant Strings of Sensors, Data Buses, and Computers, and Shut Thruster Valves and Switch to Redundant Thrusters (not necessarily in that order) (ACFD-ACRED)**
  4) **Excessive Imbalanced Thruster Commanding; Shut Thruster Valves and Switch to Redundant Thrusters, Sequenced Switch to Redundant Strings of Sensors, Data Buses, and Computers (not necessarily in that order, though thruster valve closures and switches would likely be the first action) (TCMD-THREDSW)**

Detection Coverage

- **Must define the goal that is being protected**
  - For our example, the main goal is amount of science data gathered

- **Usually assessed "from the bottom up" using the failure modes that produce the effects defined in failure scenarios**
  - Simplest Approach – non-probabilistic sum
    - Sum up the failure modes selected as above, and those potentially detected by the failure detection mechanism are divided by the total: $C = \Sigma F_{det} / \Sigma F_{tot}$
  - More complex but more insightful approach – probability-weighted sum of selected failure modes, divided by probability-weighted sum of all failure modes

- **In practice often have to use engineering estimates not directly tied to failure modes**

- **Can also do a top-down non-quantitative approach**
  - Important but not described in paper as it emphasizes quantitative estimates

- **Degradation detection coverage handled similarly (used for prognostics)**

- **Anomaly coverage is very different**
  - Anomaly = unexpected performance of intended function
  - Suggest identification of state variables associated with goals, and comparing detection mechanisms to the expected range of those variables

Example: Detection Coverage & Effectiveness

Table 2. Detection Coverage of Failure Scenarios

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed
GLC-C	99%			
GLC-O	99%			
GLC-S	99.5%			
LTC-C		100%		
LTC-S		100%		
LAC-C			90%	10%
LAC-O			100%	
LAC-S			90%	10%
PRP-C			20%	70%
PRP-S			20%	75%

Table 3: Risk Coverage and non-Covered Risk Absolute Values

	Scenario Risk	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed	Non-Covered LOM Risk
GLC-C	1E-4	9.9E-5				1E-6
GLC-O	1E-5	9.9E-6				1E-7
GLC-S	3E-4	2.985E-4				1.5E-6
LTC-C	1E-5		1E-5			0
LTC-S	2E-5		2E-5			0
LAC-C	2E-4			1.8E-4	2E-5	0
LAC-O	3E-4			3E-4		0
LAC-S	5E-4			4.5E-4	5E-5	0
PRP-C	3E-4			6E-5	2.1E-4	3E-5
PRP-S	4E-4			8E-5	3E-4	2E-5

Table 4. Detection Effectiveness per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed
GLC-C	99.9%			
GLC-O	95%			
GLC-S	99.9%			
LTC-C		100%		
LTC-S		100%		
LAC-C			99%	99%
LAC-O			98%	
LAC-S			99.5%	97.5%
PRP-C			99%	99%
PRP-S			99.5%	97.5%

Table 5. Detection Effectiveness Cumulative Risks per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed	Cumulative LOM Risk
GLC-C	9.89E-5				1.099E-6
GLC-O	9.405 E-6				5.95E-7
GLC-S	2.984E-4				1.649E-6
LTC-C		1E-5			0
LTC-S		2E-5			0
LAC-C			1.782E-4	1.98E-5	2E-6
LAC-O			2.94E-4		6E-6
LAC-S			4.478E-4	4.875E-5	3.5E-6
PRP-C			5.94E-5	2.079E-4	3.27E-5
PRP-S			7.96E-5	2.925E-4	2.79E-5

Fault Diagnostics

- **Diagnostics include Fault Isolation & Identification**
  - Isolation = Determining the possible locations of a hypothesized failure or anomaly cause, to a defined level of granularity. *(note this differs from the EE usage)*
  - Identification = Determining the possible causes of a failure or anomaly.

- **Both essentially the same in methodology; isolation determines the location of the cause, and identification determines the cause.**
  - Example: There are 3 components, each with 5 failure modes. A failure has been detected. Isolation determines which of the 3 components the cause resides in, and identification identifies one of the 5 failure modes inside that component as the cause.

- **Typically, the metrics must deal with "ambiguity groups"**
  - Example: In the example above, isolation may only determine that the cause is in components A and B, but not C. Identification may then identify that of the 10 failure modes in components A and B, 7 of the 10 failure modes are possible. In each case, the 2 components and the 7 failure modes are considered "ambiguity groups" in which you cannot distinguish between locations or failure modes, respectively.

- **Effectiveness estimates depend on the set of responses available, and how many can be taken for an "acceptable result".**
  - Example: for example above, it is acceptable to have an ambiguity group of 2 or 3 if the responses properly address the 2 or 3 possible locations/causes. This frequently means taking a sequence of actions until the problem is resolved.

Prognostics

- **Prognostics = predicting the time at which a component will no longer perform its intended function.**

- **Primary output of prognostics is “Remaining Useful Life” (RUL), which is the time from the present to the time the component fails (no longer performs intended function).**
  - The purpose of prognostics and RUL is to provide information that guides proper response, which consists usually of repair, replacement, retirement, or other mitigations to extend RUL.

- **RUL uses TPMs for accuracy, precision, and convergence.**
  - Accuracy = closeness of predictive estimate to actual value
  - Precision = variability of prediction
  - Convergence = quantification of how accuracy and precision measures improve as RUL decreases (improvement over time before failure occurs)
  - These all relate to uncertainties of estimates.

- **Truth Table methods apply (TP/TN/FP/FN)**
  - An RUL that is too long in time can lead to system failure, which is an FN.
  - An RUL that is too short in time or projects a failure that does not occur (TTC in reality is “infinity” creates a False Positive.

Failure Response Determination
(Decision Function)

- Failure Response Determination = Selecting actions to mitigate a current or future failure.

- Ultimate metric is the probability that the correct action(s) was/is/will be selected.

- This is a research topic---I know of no existing research on how to perform such estimates, but incorrect response decisions can definitely be made.

- Incorrect response decisions more likely with rushed human decisions in stressful circumstances, for example.

Example: Detection Coverage & Effectiveness

Described in Text. Isolation Effectiveness per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed
GLC-C	100%			
GLC-O	100%			
GLC-S	100%			
LTC-C		100%		
LTC-S		100%		
LAC-C			100%	100%
LAC-O			100%	
LAC-S			99%	99%
PRP-C			100%	100%
PRP-S			99%	99%

Table 6: Isolation Cumulative Risks per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed	Cumulative LOM Risk
GLC-C	9.89E-5				1.099E-6
GLC-O	9.405E-6				5.95E-7
GLC-S	2.984E-4				1.649E-6
LTC-C		1E-5			0
LTC-S		2E-5			0
LAC-C			1.782E-4	1.98E-5	2E-6
LAC-O			2.94E-4		6E-6
LAC-S			4.433E-4	4.826E-5	8.465E-6
PRP-C			5.94E-5	2.079E-4	3.27E-5
PRP-S			7.88E-5	2.89E-4	3.162E-5

Described in Text. Response Decision Effectiveness per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed
GLC-C	100%			
GLC-O	100%			
GLC-S	100%			
LTC-C		100%		
LTC-S		99%		
LAC-C			100%	100%
LAC-O			100%	
LAC-S			100%	100%
PRP-C			100%	100%
PRP-S			100%	100%

Table 7. Response Decision Cumulative Risks per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed	Cumulative LOM Risk
GLC-C	9.89E-5				1.099E-6
GLC-O	9.405E-6				5.95E-7
GLC-S	2.984E-4				1.649E-6
LTC-C		1E-5			0
LTC-S		1.98E-5			2E-7
LAC-C			1.782E-4	1.98E-5	2E-6
LAC-O			2.94E-4		6E-6
LAC-S			4.433E-4	4.826E-5	8.465E-6
PRP-C			5.94E-5	2.079E-4	3.27E-5
PRP-S			7.88E-5	2.896E-4	3.162E-5

Response Effectiveness & Total Value

- **All responses must operate more quickly than the failure effects that they are mitigating / responding to**
  - Therefore a race condition analysis of failure effects versus failure responses is required.
  - The race includes latencies for detection, diagnostics (isolation/identification), decision, and response, versus failure effect propagations to the “Critical Failure Effect”.
- **Responses may also have design faults, or if performed by humans, human faults**
  - Most likely cause of response failure, aside from losing the race condition, is interactions between failure responses, and between responses and other system control activities
- **Once response effectiveness is estimated and residual risks calculated for all scenarios, total value of the FMCLs are calculated by summing the total risk reduction / benefit across all scenarios.**
  - Can calculate effectiveness of each FMCL as a fraction of risk that it mitigates.
- **Can then estimate the total value of the FM as a whole for the system by summing the value of all FMCL benefit values.**

Example: Response Effectiveness and Total FMCL Value

Table 8. Response Effectiveness per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed
GLC-C	95%			
GLC-O	0%			
GLC-S	92%			
LTC-C		96%		
LTC-S		98%		
LAC-C			100%	98%
LAC-O			90%	
LAC-S			99%	95%
PRP-C			98%	98%
PRP-S			98%	95%

Table 9: FMCL Cumulative Risks per Failure Scenario

	WT-RbtSf	Trnd-Att&Pwr	ACFD-ACRed	TCMD-THRed	Cumulative LOM Risk
GLC-C	9.396E-5				6.044E-6
GLC-O	0				1E-5
GLC-S	2.745E-4				2.522E-5
LTC-C		9.6E-6			4E-7
LTC-S		1.94E-5			5.96E-7
LAC-C			1.782E-4	1.94E-5	2.396E-6
LAC-O			2.646E-4		3.54E-5
LAC-S			4.388E-4	4.585E-5	1.531E-5
PRP-C			5.821E-5	2.037E-4	3.805E-5
PRP-S			7.723E-5	2.751E-4	4.768E-5

Table 10. FMCL Value Estimates

FMCL Name	Original Risk	Non-Covered Risk	Covered Risk	LOS/LOM Benefit	Effectiveness Fraction
WT-RbtSf	4.10E-4	2.60E-6	4.07E-4	3.68E-4	90.4%
Trnd-Att&Pwr	3.00E-5	0	3.00E-5	2.90E-5	96.7%
ACFD-ACRed	1.07E-3	0	1.07E-3	1.02E-3	95.1%
TCMD-THRed	6.30E-4	5.00E-5	5.80E-4	5.44E-4	93.8%

Conclusion

- **FM Metrics are based on a controls-based theory of FM.**
- **FM Metrics are calculated for each FMCL for each Failure Scenario**
- **Analyze the effectiveness of each FMCL, by estimating the effectiveness of each FM function in the FMCL (detection, diagnostics, prognostics, decision, response), and cumulatively accumulating risk (or deceasing benefit).**
- **Estimate the value of each FMCL by summing the values of those FMCLs across all Failure Scenarios.**
  - This needs to be balanced against the cost of each FMCL in terms of resources, schedule, and added false positive risk.
- **Total value of FM estimated by summing the value of each FMCL.**